A Unifying Analysis of Projected Gradient Descent for ℓ_p-constrained Least Squares

S. Bahmani[a,*], B. Raj[a,b]

[a]*Department of Electrical & Computer Engineering, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213*
[b]*Language Technologies Institute, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213*

Abstract

In this paper we study the performance of the Projected Gradient Descent (PGD) algorithm for ℓ_p-constrained least squares problems that arise in the framework of Compressed Sensing. Relying on the Restricted Isometry Property, we provide convergence guarantees for this algorithm for the entire range of $0 \leq p \leq 1$, that include and generalize the existing results for the Iterative Hard Thresholding algorithm and provide a new accuracy guarantee for the Iterative Soft Thresholding algorithm as special cases. Our results suggest that in this group of algorithms, as p increases from zero to one, conditions required to guarantee accuracy become stricter and robustness to noise deteriorates.

Keywords: Least Squares, Compressed Sensing, Sparsity, Underdetermined Linear Systems, Restricted Isometry Property, Projected Gradient Descent

1. Introduction

Least squares problems occur in various signal processing and statistical inference applications. In these problems the relation between the vector of noisy observations $\mathbf{y} \in \mathbb{C}^m$ and the unknown parameter or signal $\mathbf{x}^\star \in \mathbb{C}^n$ is governed by a linear equation of the form

$$\mathbf{y} = \mathbf{A}\mathbf{x}^\star + \mathbf{e}, \tag{1}$$

where $\mathbf{A} \in \mathbb{C}^{m \times n}$ is a matrix that may model a linear system or simply contains a set of collected data. The vector $\mathbf{e} \in \mathbb{C}^m$ represents the additive observation noise. Estimating $\mathbf{x}^\star$ from the observation vector $\mathbf{y}$ is achieved by finding the $\mathbf{x} \in \mathbb{C}^n$ that minimizes the squared error $\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2$. This least squares approach, however, is well-posed only if the nullspace of matrix $\mathbf{A}$ merely contains the zero vector. The cases in which the nullspace is greater than the singleton $\{\mathbf{0}\}$, as in underdetermined scenarios ($m < n$), are more relevant in a variety of applications. To enforce unique least squares solutions in these cases, it becomes necessary to have some prior information about the structure of $\mathbf{x}^\star$.

One of the structural characteristics that describes parameters and signals of interest in a wide range of applications from medical imaging to astronomy is *sparsity*. Since the advent of the theory of *compressed sensing*, development and analysis of algorithms that exploit sparsity for estimation in underdetermined problems have become important topics of study. In the absence of noise $\mathbf{x}^\star$ can be uniquely determined from the observation vector $\mathbf{y} = \mathbf{A}\mathbf{x}^\star$, provided that spark$(\mathbf{A}) > 2\|\mathbf{x}^\star\|_0$ (i.e., every $2\|\mathbf{x}^\star\|_0$ columns of $\mathbf{A}$ are linearly independent) [12]. Then the ideal estimation procedure could simply be finding the sparsest vector $\mathbf{x}$ that incurs no residual error (i.e., $\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 = 0$). This ideal estimation method can be extended

---

to the case of noisy observations as well. Formally, given an upper bound ϵ on the ℓ_2-norm of the noise, the vector $\mathbf{x}^\star$ can be estimated by solving the ℓ_0-minimization

$$\arg\min_{\mathbf{x}} \ \|\mathbf{x}\|_0 \quad \text{s.t.} \ \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon, \tag{2}$$

where $\|\mathbf{x}\|_0$ denotes the ℓ_0-norm[1] of the vector $\mathbf{x}$ that merely counts the number of its non-zero entries. However, this minimization problem is in general NP-hard [17]. To avoid the combinatorial computational cost of (2), often the ℓ_0-norm is substituted by the ℓ_p-norm[1] $\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$ for some $p \in (0,1]$ providing the ℓ_p-minimization

$$\arg\min_{\mathbf{x}} \ \|\mathbf{x}\|_p \quad \text{s.t.} \ \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon. \tag{3}$$

In particular, at $p = 1$ the ℓ_1-minimization can be solved in polynomial time using convex programming algorithms. Several theoretical and experimental results [see e.g., 7, 20, 21] suggest that ℓ_p-minimization with $p \in (0,1)$ requires fewer observations than the ℓ_1-minimization to produce accurate estimates. However, ℓ_p-minimization is a non-convex problem where finding the global minimizer is not guaranteed and can be computationally more expensive than the ℓ_1-minimization.

An alternative approach in the framework of sparse linear regression is to solve the sparsity-constrained least squares problem

$$\arg\min_{\mathbf{x}} \ \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \quad \text{s.t.} \ \|\mathbf{x}\|_0 \leq s, \tag{4}$$

where $s = \|\mathbf{x}^\star\|_0$ is given. Similar to (2) solving (4) is not tractable and approximate solvers must be sought. Several compressed sensing algorithms jointly known as the *greedy pursuits* including Iterative Hard Thresholding (IHT) [3], Subspace Pursuit (SP) [10], and Compressive Sampling Matching Pursuit (CoSaMP) [18] are implicitly approximate solvers of (4).

As a relaxation of (4) one may also consider the ℓ_p-constrained least squares

$$\arg\min_{\mathbf{x}} \ \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \quad \text{s.t.} \ \|\mathbf{x}\|_p \leq R^\star, \tag{5}$$

given $R^\star = \|\mathbf{x}^\star\|_p$. The Least Absolute Shrinkage and Selection Operator (LASSO) [22] is a well-known special case of this optimization problem with $p = 1$. The optimization problem of (5) typically does not have a closed-form solution, but can be (approximately) solved using iterative Projected Gradient Descent (PGD), which has been outlined in Section 2. Previous studies of these algorithms, henceforth referred to as ℓ_p-PGD, are limited to the cases of $p = 0$ and $p = 1$. The algorithm corresponding to the case of $p = 0$ is recognized in the literature as the IHT algorithm. The Iterative Soft Thresholding (IST) algorithm [2] is originally proposed as a solver of the Basis Pursuit Denoising (BPDN) [9], which is the unconstrained equivalent of the LASSO with the ℓ_1-norm as the regularization term. However, the IST algorithm also naturally describes a PGD solver of (5) for $p = 1$ [see for e.g, 1] by considering varying shrinkage in iterations, as described in [2], to enforce the iterates to have sufficiently small ℓ_1-norm. The main contribution of this paper is a comprehensive analysis of the performance of ℓ_p-PGD algorithms for the entire regime of $p \in [0,1]$.

In the extreme case of $p = 0$ we have the ℓ_0-PGD algorithm which is indeed the IHT algorithm. Unlike conventional PGD algorithms, the feasible set —the set of points that satisfy the optimization constraints— for IHT is the non-convex set of s-sparse vectors. Therefore, the standard analysis for PGD algorithms with convex feasible sets that relies on the fact that projection onto convex sets defines a contraction map will no longer apply. However, imposing extra conditions on the matrix $\mathbf{A}$ can be leveraged to provide convergence guarantees [3, 13].

At $p = 1$ where (5) is a convex program, the corresponding ℓ_1-PGD algorithm has been studied under the name of IST in different scenarios (see [2] and references therein). Ignoring the sparsity of the vector $\mathbf{x}^\star$,

---

[1]The term "norm" is used for convenience throughout the paper. In fact, the ℓ_0 functional violates the positive scalability property of the norms and the ℓ_p functionals with $p \in (0,1)$ are merely *quasi-norms*.

it can be shown that the IST algorithm exhibits a sublinear rate of convergence as a convex optimization algorithm [2]. In the context of the *sparse* estimation problems, however, faster rates of convergence can be guaranteed for IST. For example, in [1] PGD algorithms are studied in a broad category of regression problems regularized with "decomposable" norms. In this configuration, which includes sparse linear regression via IST, the PGD algorithms are shown to possess a linear rate of convergence provided the objective function—the squared error in our case—satisfies *Restricted Strong Convexity* (RSC) and *Restricted Smoothness* (RSM) conditions [1]. These two conditions basically control the curvature of the objective function being restricted to (nearly) sparse vectors. Although the results provided in [1] consolidate the analysis of several interesting problems, they do not readily extend to the case of ℓ_p-constrained least squares since the constraint is not defined by a true norm.

In this paper, by considering ℓ_p-balls of given radii as feasible sets in the general case, we study the ℓ_p-PGD algorithms that render a continuum of sparse reconstruction algorithms, and encompass both the IHT and the IST algorithms. In Section 2 using the Restricted Isometry Property (RIP) [5] we provide accuracy guarantees for ℓ_p-PGD algorithms which assert that these algorithms converge to the true signal up to a multiple of the noise level at a linear rate. Furthermore, our results suggest that as p increases from zero to one the convergence and robustness to noise deteriorates. This conclusion is particularly in agreement with the empirical studies of the *phase transition* of the IST and IHT algorithms provided in [16]. Our results for ℓ_0-PGD coincides with the guarantees for IHT derived in [13]. Furthermore, to the best of our knowledge the RIP-based accuracy guarantees we provide for IST, which is the ℓ_1-PGD algorithm, have not been derived before. The last section of the paper, Section 3, is dedicated to discussion of some details and future work.

Notation. Throughout the paper we assume that the vectors and matrices have complex entries unless stated otherwise. The set $\{1, 2, \ldots, n\}$ is denoted by $[n]$ for brevity. We use $\mathbf{M}_{\mathcal{J}}$ to denote restriction of the matrix $\mathbf{M}$ to the columns selected by the set of indices $\mathcal{J} \subseteq [n]$. Similarly, $\mathbf{v}|_{\mathcal{J}}$ denotes restriction of the vector $\mathbf{v}$ to the entries with indices in $\mathcal{J}$. Depending on the context, the vector $\mathbf{v}|_{\mathcal{J}}$ may also denote a vector that is equal to the vector $\mathbf{v}$ except for the part supported on $\mathcal{J}^{\mathsf{c}}$ where it is zero. The set of non-zero entries (i.e, the support set) and the best s-term approximation of vector $\mathbf{v}$ are denoted by $\mathrm{supp}\,(\mathbf{v})$ and $\mathbf{v}_s$, respectively. Furthermore, the matrix $\mathbf{M}^{\mathsf{H}}$ denotes the Hermitian conjugate of the matrix $\mathbf{M}$. The inner product of vectors $\mathbf{u}$ and $\mathbf{v}$ is denoted by $\langle \mathbf{u}, \mathbf{v} \rangle$. Finally, $\Re\,[\cdot]$ and $\mathrm{Arg}\,(\cdot)$ denote the real part and the phase of their arguments, respectively.

2. Projected Gradient Descent for ℓ_p-constrained Least Squares

One of the most elementary tools in convex optimization for constrained minimization is the PGD method. For a differentiable convex objective function $f\,(\cdot)$, a convex set $\mathcal{Q}$, and a projection operator $\mathrm{P}_{\mathcal{Q}}\,(\cdot)$ defined by

$$\mathrm{P}_{\mathcal{Q}}\,(\mathbf{x}) = \arg\min_{\mathbf{u}}\; \|\mathbf{x} - \mathbf{u}\|_2^2 \quad \text{s.t. } \mathbf{u} \in \mathcal{Q}, \tag{6}$$

Algorithm 1: Project Gradient Descent

input : Objective function $f\,(\cdot)$ and an operator $\mathrm{P}_{\mathcal{Q}}\,(\cdot)$ that performs projection onto the set $\mathcal{Q}$

Choose the initial point $\mathbf{x}^0 \in \mathcal{Q}$

$k \longleftarrow 0$

repeat

    Choose a step-size $\eta_k > 0$

    $\mathbf{x}^{k+1} \longleftarrow \mathrm{P}_{\mathcal{Q}}\,(\mathbf{x}^k - \eta_k \nabla f\,(\mathbf{x}^k))$

    $k \longleftarrow k + 1$

until *halting condition holds*

output: the (approximate) minimizer $\mathbf{x}^k$

---

the PGD algorithm solves the minimization

$$\arg\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{s.t. } \mathbf{x} \in \mathcal{Q}$$

via the iterations outlined in Algorithm 1. For example, in a broad range of applications where the objective function is the squared error of the form $f(\mathbf{x}) = \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2$, the iterate update equation of the PGD method in Algorithm 1 reduces to

$$\mathbf{x}^{k+1} = \mathrm{P}_{\mathcal{Q}}\left(\mathbf{x}^k - \eta_k \mathbf{A}^{\mathsf{H}}\left(\mathbf{A}\mathbf{x}^k - \mathbf{y}\right)\right). \tag{7}$$

In the context of compressed sensing if (1) holds and $\mathcal{Q}$ is the ℓ_1-ball of radius $\|\mathbf{x}^\star\|_1$ centered at the origin, Algorithm 1 reduces to the IST algorithm (except perhaps for variable step-size) that solves (5) for $p = 1$. By relaxing the convexity restriction imposed on $\mathcal{Q}$ the PGD iterations also describe the IHT algorithm where $\mathcal{Q}$ is the set of vectors whose ℓ_0-norm is not greater than $s = \|\mathbf{x}^\star\|_0$.

Henceforth, we refer to an ℓ_p-ball centered at the origin and aligned with the axes simply as an ℓ_p-ball for brevity. To proceed let us define the set

$$\mathcal{F}_p(c) = \left\{\mathbf{x} \in \mathbb{C}^n \mid \sum_{i=1}^{n} |x_i|^p \le c\right\}, \tag{8}$$

for $c \in \mathbb{R}^+$, which describes an ℓ_p-ball. Although c can be considered as the radius of this ℓ_p-ball with respect to the metric $d(\mathbf{a}, \mathbf{b}) = \|\mathbf{a} - \mathbf{b}\|_p^p$, we call c the "p-radius" of the ℓ_p-ball to avoid confusion with the conventional definition of the radius for an ℓ_p-ball, i.e., $\max_{\mathbf{x} \in \mathcal{F}_p(c)} \|\mathbf{x}\|_p$. Furthermore, at $p = 0$ where $\mathcal{F}_p(c)$ describes the same "ℓ_0-ball" different values of c, we choose the smallest c as the p-radius of the ℓ_p-ball for uniqueness. In this section we will show that to estimate the signal $\mathbf{x}^\star$ that is either sparse or *compressible* in fact the PGD method can be applied in a more general framework where the feasible set is considered to be an ℓ_p-ball of given p-radius. Ideally the p-radius of the feasible set should be $\|\mathbf{x}^\star\|_p^p$, but in practice this information might not be available. In our analysis, we merely assume that the p-radius of the feasible set is not greater than $\|\mathbf{x}^\star\|_p^p$, i.e., the feasible set does not contain $\mathbf{x}^\star$ in its interior.

Note that for the feasible sets $\mathcal{Q} = \mathcal{F}_p(c)$ with $p \in (0, 1]$ the minimum value in (6) is always attained because the objective is continuous and the set $\mathcal{Q}$ is compact. Therefore, there is at least one minimizer in $\mathcal{Q}$. However, for $p < 1$ the set $\mathcal{Q}$ is nonconvex and there might be multiple projection points in general. For the purpose of the analysis presented in this paper, however, any such minimizer is acceptable. Using the axiom of choice, we can assume existence of a choice function that for every $\mathbf{x}$ selects one of the solutions of (6). This function indeed determines a projection operator which we denote by $\mathrm{P}_{\mathcal{Q}}(\mathbf{x})$.

Many compressed sensing algorithms such as those of [3, 4, 10, 18] rely on sufficient conditions expressed in terms of the RIP of the matrix $\mathbf{A}$. We also provide accuracy guarantees of the ℓ_p-PGD algorithm with the assumption that certain RIP conditions hold. The following definition states the RIP in its asymmetric form. This definition is previously proposed in the literature [14], though in a slightly different format.

Definition (RIP). Matrix $\mathbf{A}$ is said to have RIP of order s with restricted isometry constants α_s and β_s if they are in order the smallest and the largest non-negative numbers such that

$$\beta_s \|\mathbf{x}\|_2^2 \le \|\mathbf{A}\mathbf{x}\|_2^2 \le \alpha_s \|\mathbf{x}\|_2^2$$

holds for all s-sparse vectors $\mathbf{x}$.

In the literature usually the symmetric form of the RIP is considered in which $\alpha_s = 1 + \delta_s$ and $\beta_s = 1 - \delta_s$ with $\delta_s \in [0, 1]$. For example, in [13] the ℓ_1-minimization is shown to accurately estimate $\mathbf{x}^\star$ provided $\delta_{2s} < 3/(4 + \sqrt{6}) \approx 0.46515$. Similarly, accuracy of the estimates obtained by IHT, SP, and CoSaMP are guaranteed provided $\delta_{3s} < 1/2$ [13], $\delta_{3s} < 0.205$ [10], and $\delta_{4s} < \sqrt{2/(5 + \sqrt{73})} \approx 0.38427$ [13], respectively.

As our first contribution, in the following theorem we show that the ℓ_p-PGD accurately solves ℓ_p-constrained least squares provided the matrix $\mathbf{A}$ satisfies a proper RIP criterion. To proceed we define

$$\rho_s = \frac{\alpha_s - \beta_s}{\alpha_s + \beta_s},$$

4

which can be interpreted as the equivalent of the standard RIP constant δ_s in the asymmetric form of RIP.

Theorem 2.1. *Let $\mathbf{x}^\star$ be an s-sparse vector whose compressive measurements are observed according to* (1) *using a measurement matrix $\mathbf{A}$ that satisfies RIP of order $3s$. To estimate $\mathbf{x}^\star$ via the ℓ_p-PGD algorithm an ℓ_p-ball $\widehat{\mathcal{B}}$ with p-radius $\widehat{c}$ (i.e., $\widehat{\mathcal{B}} = \mathcal{F}_p(\widehat{c})$) is given as the feasible set for the algorithm such that $\widehat{c} = (1 - \varepsilon)^p \|\mathbf{x}^\star\|_p^p$ for some[2] $\varepsilon \in [0, 1)$. Furthermore, suppose that the step-size η_k of the algorithm can be chosen to obey $\left| \frac{\eta_k (\alpha_{3s} + \beta_{3s})}{2} - 1 \right| \leq \tau$ for some $\tau \geq 0$. If*

$$(1 + \tau) \rho_{3s} + \tau < \frac{1}{2 \left(1 + \sqrt{2} \xi(p)\right)^2} \tag{9}$$

with $\xi(p)$ denoting the function $\sqrt{p} \left(\frac{2}{2-p}\right)^{1/2 - 1/p}$, then $\mathbf{x}^k$, the k-th iterate of the algorithm, obeys

$$\left\|\mathbf{x}^k - \mathbf{x}^\star\right\|_2 \leq (2\gamma)^k \|\mathbf{x}^\star\|_2 + \frac{2(1 + \tau)}{1 - 2\gamma} (1 + \xi(p)) \left(\varepsilon (1 + \rho_{3s}) \|\mathbf{x}^\star\|_2 + \frac{2\sqrt{\alpha_{2s}}}{\alpha_{3s} + \beta_{3s}} \|\mathbf{e}\|_2 \right) + \varepsilon \|\mathbf{x}^\star\|_2, \tag{10}$$

where

$$\gamma = ((1 + \tau) \rho_{3s} + \tau) \left(1 + \sqrt{2} \xi(p)\right)^2. \tag{11}$$

Remark 2.1. Note that the parameter ε indicates how well the feasible set $\widehat{\mathcal{B}}$ approximates the ideal feasible set $\mathcal{B}^\star = \mathcal{F}_p\left(\|\mathbf{x}^\star\|_p^p\right)$. The terms in (10) that depend on ε determine the error caused by the mismatch between $\widehat{\mathcal{B}}$ and $\mathcal{B}^\star$. Ideally, one has $\varepsilon = 0$ and the residual error becomes merely dependent on the noise level $\|\mathbf{e}\|_2$.

Remark 2.2. The parameter τ determines the deviation of the step-size η_k from $\frac{2}{\alpha_{3s} + \beta_{3s}}$ which might not be known *a priori*. In this formulation, smaller values of τ are desirable since they impose less restrictive condition on ρ_{3s} and also result in smaller residual error. Furthermore, we can naively choose $\eta_k = \|\mathbf{A}\mathbf{x}\|_2^2 / \|\mathbf{x}\|_2^2$ for some $3s$-sparse vector $\mathbf{x} \neq \mathbf{0}$ to ensure $1/\alpha_{3s} \leq \eta_k \leq 1/\beta_{3s}$ and thus $\left| \eta_k \frac{\alpha_{3s} + \beta_{3s}}{2} - 1 \right| \leq \frac{\alpha_{3s} - \beta_{3s}}{2\beta_{3s}}$. Therefore, we can always assume that $\tau \leq \frac{\alpha_{3s} - \beta_{3s}}{2\beta_{3s}}$.

Remark 2.3. Note that the function $\xi(p)$, depicted in Fig. 1, controls the variation of the stringency of the condition (9) and the variation of the residual error in (10) in terms of p. Straightforward algebra shows that $\xi(p)$ is an increasing function of p with $\xi(0) = 0$. Therefore, as p increases from zero to one, the RHS of (9) decreases, which implies the measurement matrix must have a smaller ρ_{3s} to satisfy the sufficient condition (9). Similarly, as p increases from zero to one the residual error in (10) increases. To contrast this result with the existing guarantees of other iterative algorithms, suppose that $\tau = 0$, $\varepsilon = 0$, and we use the symmetric form of RIP (i.e., $\alpha_{3s} = 1 + \delta_{3s}$ and $\beta_{3s} = 1 - \delta_{3s}$) which implies $\rho_{3s} = \delta_{3s}$. At $p = 0$, corresponding to the IHT algorithm, (9) reduces to $\delta_{3s} < 1/2$ that is identical to the condition derived in [13]. Furthermore, the required condition at $p = 1$, corresponding to the IST algorithm, would be $\delta_{3s} < 1/8$.

The guarantees stated in Theorem 2.1 can be generalized for nearly sparse or *compressible* signals that can be defined using power laws as described in [6]. The following corollary provides error bounds for a general choice of $\mathbf{x}^\star$.

Corollary 2.1. *Suppose that $\mathbf{x}^\star$ is an arbitrary vector in $\mathbb{C}^n$ and the conditions of Theorem 2.1 hold for $\mathbf{x}_s^\star$, then the k-th iterate of the ℓ_p-PGD algorithm provides an estimate of $\mathbf{x}_s^\star$ that obeys*

---

[2]At $p = 0$ we have $(1 - \varepsilon)^0 = 1$ which enforces $\widehat{c} = \|\mathbf{x}^\star\|_0$. In this case ε is not unique, but to make a coherent statement we assume that $\varepsilon = 0$.
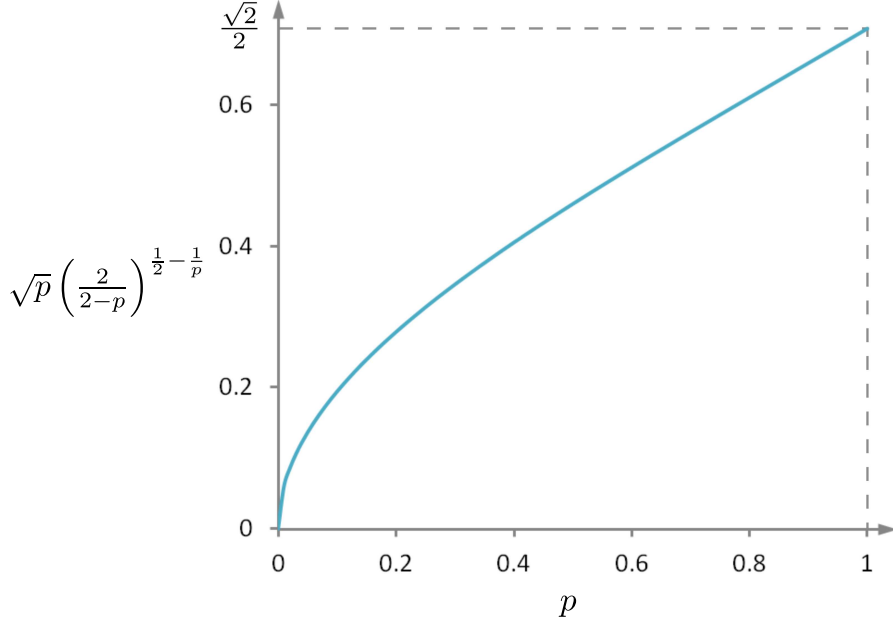
Figure 1: Plot of the function $\xi\left(p\right) = \sqrt{p}\left(\frac{2}{2-p}\right)^{\frac{1}{2}-\frac{1}{p}}$ which determines the contraction factor and the residual error.

$$\left\|\mathbf{x}^k - \mathbf{x}^\star\right\|_2 \leq (2\gamma)^k \left\|\mathbf{x}_s^\star\right\|_2 + \frac{2\left(1+\tau\right)\left(1+\xi\left(p\right)\right)}{1-2\gamma}\left(\varepsilon\left(1+\rho_{3s}\right)\left\|\mathbf{x}_s^\star\right\|_2 + \frac{2\alpha_{2s}}{\alpha_{3s}+\beta_{3s}}\left(\left\|\mathbf{x}^\star - \mathbf{x}_s^\star\right\|_2 + \frac{\left\|\mathbf{x}^\star - \mathbf{x}_s^\star\right\|_1}{\sqrt{2s}}\right)\right.$$
$$\left. + \frac{2\sqrt{\alpha_{2s}}}{\alpha_{3s}+\beta_{3s}}\left\|\mathbf{e}\right\|_2\right) + \varepsilon\left\|\mathbf{x}_s^\star\right\|_2 + \left\|\mathbf{x}^\star - \mathbf{x}_s^\star\right\|_2 .$$

Proof. Let $\widetilde{\mathbf{e}} = \mathbf{A}\left(\mathbf{x}^\star - \mathbf{x}_s^\star\right) + \mathbf{e}$. We can write $\mathbf{y} = \mathbf{A}\mathbf{x}^\star + \mathbf{e} = \mathbf{A}\mathbf{x}_s^\star + \widetilde{\mathbf{e}}$. Thus, we can apply Theorem 2.1 considering $\mathbf{x}_s^\star$ as the signal of interest and $\widetilde{\mathbf{e}}$ as the noise vector and obtain

$$\left\|\mathbf{x}^k - \mathbf{x}_s^\star\right\|_2 \leq (2\gamma)^k \left\|\mathbf{x}_s^\star\right\|_2 + \frac{2\left(1+\tau\right)}{1-2\gamma}\left(1+\xi\left(p\right)\right)\left(\varepsilon\left(1+\rho_{3s}\right)\left\|\mathbf{x}_s^\star\right\|_2 + \frac{2\sqrt{\alpha_{2s}}}{\alpha_{3s}+\beta_{3s}}\left\|\widetilde{\mathbf{e}}\right\|_2\right) + \varepsilon\left\|\mathbf{x}_s^\star\right\|_2 . \quad (12)$$

Furthermore, we have

$$\left\|\widetilde{\mathbf{e}}\right\|_2 = \left\|\mathbf{A}\left(\mathbf{x}^\star - \mathbf{x}_s^\star\right) + \mathbf{e}\right\|_2$$
$$\leq \left\|\mathbf{A}\left(\mathbf{x}^\star - \mathbf{x}_s^\star\right)\right\|_2 + \left\|\mathbf{e}\right\|_2 .$$

Then applying Proposition 3.5 of [18] yields

$$\left\|\widetilde{\mathbf{e}}\right\|_2 \leq \sqrt{\alpha_{2s}}\left(\left\|\mathbf{x}^\star - \mathbf{x}_s^\star\right\|_2 + \frac{1}{\sqrt{2s}}\left\|\mathbf{x}^\star - \mathbf{x}_s^\star\right\|_1\right) + \left\|\mathbf{e}\right\|_2 .$$

Applying this inequality in (12) followed by the triangle inequality $\left\|\mathbf{x}^k - \mathbf{x}^\star\right\|_2 \leq \left\|\mathbf{x}^k - \mathbf{x}_s^\star\right\|_2 + \left\|\mathbf{x}^\star - \mathbf{x}_s^\star\right\|_2$ yields the desired inequality. ∎

To prove Theorem 2.1 first a series of lemmas should be established. In what follows, $\mathbf{x}_\perp^\star$ is a projection of the s-sparse vector $\mathbf{x}^\star$ onto $\widehat{\mathcal{B}}$ and $\mathbf{x}^\star - \mathbf{x}_\perp^\star$ is denoted by $\mathbf{d}^\star$. Furthermore, for $k = 0, 1, 2, \ldots$ we denote $\mathbf{x}^k - \mathbf{x}_\perp^\star$ by $\mathbf{d}^k$ for compactness.

Lemma 2.1. *If $\mathbf{x}^k$ denotes the estimate in the k-th iteration of ℓ_p-PGD, then*

$$\left\|\mathbf{d}^{k+1}\right\|_2^2 \le 2\Re\left[\left\langle\mathbf{d}^k, \mathbf{d}^{k+1}\right\rangle - \eta_k\left\langle\mathbf{A}\mathbf{d}^k, \mathbf{A}\mathbf{d}^{k+1}\right\rangle\right] + 2\eta_k\Re\left\langle\mathbf{A}\mathbf{d}^{k+1}, \mathbf{A}\mathbf{d}^\star + \mathbf{e}\right\rangle.$$

Proof. Note that $\mathbf{x}^{k+1}$ is a projection of $\mathbf{x}^k - \eta_k\mathbf{A}^{\mathsf{H}}\left(\mathbf{A}\mathbf{x}^k - \mathbf{y}\right)$ onto $\widehat{\mathcal{B}}$. Since $\mathbf{x}_\perp^\star$ is also a feasible point (i.e., $\mathbf{x}_\perp^\star \in \widehat{\mathcal{B}}$) we have

$$\left\|\mathbf{x}^{k+1} - \mathbf{x}^k + \eta_k\mathbf{A}^{\mathsf{H}}\left(\mathbf{A}\mathbf{x}^k - \mathbf{y}\right)\right\|_2^2 \le \left\|\mathbf{x}_\perp^\star - \mathbf{x}^k + \eta_k\mathbf{A}^{\mathsf{H}}\left(\mathbf{A}\mathbf{x}^k - \mathbf{y}\right)\right\|_2^2.$$

Using (1) we obtain

$$\left\|\mathbf{d}^{k+1} - \mathbf{d}^k + \eta_k\mathbf{A}^{\mathsf{H}}\left(\mathbf{A}\left(\mathbf{d}^k - \mathbf{d}^\star\right) - \mathbf{e}\right)\right\|_2^2 \le \left\|-\mathbf{d}^k + \eta_k\mathbf{A}^{\mathsf{H}}\left(\mathbf{A}\left(\mathbf{d}^k - \mathbf{d}^\star\right) - \mathbf{e}\right)\right\|_2^2.$$

Therefore, we obtain

$$\Re\left\langle\mathbf{d}^{k+1}, \mathbf{d}^{k+1} - 2\mathbf{d}^k + 2\eta_k\mathbf{A}^{\mathsf{H}}\left(\mathbf{A}\mathbf{d}^k - \left(\mathbf{A}\mathbf{d}^\star + \mathbf{e}\right)\right)\right\rangle \le 0$$

that yields the the desired result after straightforward algebraic manipulations. ∎

The following lemma is a special case of the generalized shifting inequality proposed in [13, Theorem 2]. Please refer to the reference for the proof.

Lemma 2.2 (*Shifting Inequality* [13]). *If $0 < p < 2$ and*

$$u_1 \ge u_2 \ge \cdots \ge u_l \ge u_{l+1} \ge \cdots \ge u_r \ge u_{r+1} \ge \cdots \ge u_{r+l} \ge 0,$$

then for $C = \max\left\{r^{\frac{1}{2}-\frac{1}{p}}, \sqrt{\frac{p}{2}}\left(\frac{2}{2-p}l\right)^{\frac{1}{2}-\frac{1}{p}}\right\}$,

$$\left(\sum_{i=l+1}^{l+r}u_i^2\right)^{\frac{1}{2}} \le C\left(\sum_{i=1}^r u_i^p\right)^{\frac{1}{p}}. \tag{13}$$

Lemma 2.3. *For $\mathbf{x}_\perp^\star$, a projection of $\mathbf{x}^\star$ onto $\widehat{\mathcal{B}}$, we have $\mathrm{supp}\left(\mathbf{x}_\perp^\star\right) \subseteq \mathcal{S} = \mathrm{supp}\left(\mathbf{x}^\star\right)$.*

Proof. Proof is by contradiction. Suppose that there exists a coordinate i such that $x_i^\star = 0$ but $x_{\perp i}^\star \ne 0$. Then one can construct vector $\mathbf{x}'$ which is equal to $\mathbf{x}_\perp^\star$ except at the i-th coordinate where it is zero. Obviously $\mathbf{x}'$ is feasible because $\|\mathbf{x}'\|_p^p < \|\mathbf{x}_\perp^\star\|_p^p \le \widehat{c}$. Furthermore,

$$\begin{aligned}
\|\mathbf{x}^\star - \mathbf{x}'\|_2^2 &= \sum_{j=1}^n\left|x_j^\star - x_j'\right|^2 \\
&= \sum_{\substack{j=1 \\ j\ne i}}^n\left|x_j^\star - x_{\perp j}^\star\right|^2 \\
&< \sum_{j=1}^n\left|x_j^\star - x_{\perp j}^\star\right|^2 \\
&= \|\mathbf{x}^\star - \mathbf{x}_\perp^\star\|_2^2.
\end{aligned}$$

This is a contradiction since by definition

$$\mathbf{x}_\perp^\star \in \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{x}^\star - \mathbf{x}\|_2^2 \quad \text{s.t.} \ \ \|\mathbf{x}\|_p^p \le \widehat{c}.$$
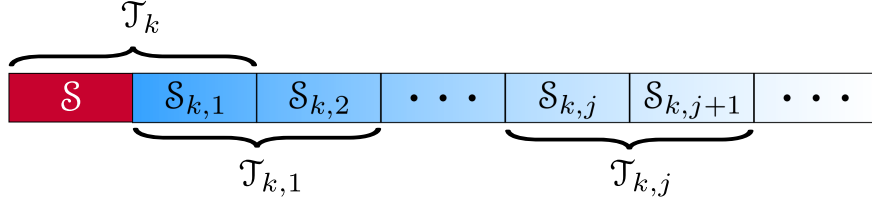
∎

Figure 2: Partitioning of vector $\mathbf{d}^k = \mathbf{x}^k - \mathbf{x}_\perp^\star$. The color gradient represents decrease of the magnitudes of the corresponding coordinates.

To continue, we introduce the following sets which partition the coordinates of vector $\mathbf{d}^k$ for $k = 0, 1, 2, \ldots$. As defined previously in Lemma 2.3, let $\mathcal{S} = \mathrm{supp}\,(\mathbf{x}^\star)$. Lemma 2.3 shows that $\mathrm{supp}\,(\mathbf{x}_\perp^\star) \subseteq \mathcal{S}$, thus we can assume that $\mathbf{x}_\perp^\star$ is s-sparse. Let $\mathcal{S}_{k,1}$ be the support of the s largest entries of $\mathbf{d}^k|_{\mathcal{S}^c}$ in magnitude, and define $\mathcal{T}_k = \mathcal{S} \cup \mathcal{S}_{k,1}$. Furthermore, let $\mathcal{S}_{k,2}$ be the support of the s largest entries of $\mathbf{d}^k|_{\mathcal{T}_k^c}$, $\mathcal{S}_{k,3}$ be the support of the next s largest entries of $\mathbf{d}^k|_{\mathcal{T}_k^c}$, and so on. We also set $\mathcal{T}_{k,j} = \mathcal{S}_{k,j} \cup \mathcal{S}_{k,j+1}$ for $j \geq 1$. This partitioning of the vector $\mathbf{d}^k$ is illustrated in Fig. 2.

Lemma 2.4. *For $k = 0, 1, 2, \ldots$ the vector $\mathbf{d}^k$ obeys*

$$\sum_{i \geq 2} \left\| \mathbf{d}^k|_{\mathcal{S}_{k,i}} \right\|_2 \leq \sqrt{2p} \left(\frac{2s}{2-p} \right)^{\frac{1}{2} - \frac{1}{p}} \left\| \mathbf{d}^k|_{\mathcal{S}^c} \right\|_p.$$

Proof. Since $\mathcal{S}_{k,j}$ and $\mathcal{S}_{k,j+1}$ are disjoint and $\mathcal{T}_{k,j} = \mathcal{S}_{k,j} \cup \mathcal{S}_{k,j+1}$ for $j \geq 1$, we have

$$\left\| \mathbf{d}^k|_{\mathcal{S}_{k,j}} \right\|_2 + \left\| \mathbf{d}^k|_{\mathcal{S}_{k,j+1}} \right\|_2 \leq \sqrt{2} \left\| \mathbf{d}^k|_{\mathcal{T}_{k,j}} \right\|_2.$$

Adding over even j's then we deduce

$$\sum_{j \geq 2} \left\| \mathbf{d}^k|_{\mathcal{S}_{k,j}} \right\|_2 \leq \sqrt{2} \sum_{i \geq 1} \left\| \mathbf{d}^k|_{\mathcal{T}_{k,2i}} \right\|_2.$$

Because of the structure of the sets $\mathcal{T}_{k,j}$, Lemma 2.2 can be applied to obtain

$$\left\| \mathbf{d}^k|_{\mathcal{T}_{k,j}} \right\|_2 \leq \sqrt{p} \left(\frac{2s}{2-p} \right)^{\frac{1}{2} - \frac{1}{p}} \left\| \mathbf{d}^k|_{\mathcal{T}_{k,j-1}} \right\|_p. \tag{14}$$

To be precise, based on Lemma 2.2 the coefficient on the RHS should be $C = \max \left\{ (2s)^{\frac{1}{2} - \frac{1}{p}}, \sqrt{\frac{p}{2}} \left(\frac{2s}{2-p} \right)^{\frac{1}{2} - \frac{1}{p}} \right\}$. For simplicity, however, we use the upper bound $C \leq \sqrt{p} \left(\frac{2s}{2-p} \right)^{\frac{1}{2} - \frac{1}{p}}$. To verify this upper bound it suffices to show that $(2s)^{\frac{1}{2} - \frac{1}{p}} \leq \sqrt{p} \left(\frac{2s}{2-p} \right)^{\frac{1}{2} - \frac{1}{p}}$ or equivalently $\phi(p) = p \log p + (2 - p) \log (2 - p) \geq 0$ for $p \in (0, 1]$. Since $\phi(\cdot)$ is a deceasing function over $(0, 1]$, it attains its minimum at $p = 1$ which means that $\phi(p) \geq \phi(1) = 0$ as desired.

Then (14) yields

$$\sum_{j \geq 2} \left\| \mathbf{d}^k|_{\mathcal{S}_{k,j}} \right\|_2 \leq \sqrt{2p} \left(\frac{2s}{2-p} \right)^{\frac{1}{2} - \frac{1}{p}} \sum_{i \geq 1} \left\| \mathbf{d}^k|_{\mathcal{T}_{k,2i-1}} \right\|_p.$$

Since $\omega_1 + \omega_2 + \cdots + \omega_l \leq (\omega_1^p + \omega_2^p + \cdots + \omega_l^p)^{\frac{1}{p}}$ holds for $\omega_1, \cdots, \omega_l \geq 0$ and $p \in (0, 1]$, we can write

$$\sum_{i \geq 1} \left\| \mathbf{d}^k|_{\mathcal{T}_{k,2i-1}} \right\|_p \leq \left(\sum_{i \geq 1} \left\| \mathbf{d}^k|_{\mathcal{T}_{k,2i-1}} \right\|_p^p \right)^{\frac{1}{p}}.$$

The desired result then follows using the fact that the sets $\mathcal{T}_{k,2i-1}$ are disjoint and $\bigcup_{i \geq 1} \mathcal{T}_{k,2i-1} = \mathcal{S}^c$. ∎

8

Proof of the following Lemma mostly relies on some common inequalities that have been used in the compressed sensing literature (see e.g., [8, Theorem 2.1] and [15, Theorem 2]) .

Lemma 2.5. *The error vector* $\mathbf{d}^k$ *satisfies* $\left\| \mathbf{d}^k|_{\mathcal{S}^c} \right\|_p \leq s^{\frac{1}{p} - \frac{1}{2}} \left\| \mathbf{d}^k|_{\mathcal{S}} \right\|_2$ *for all* $k = 0, 1, 2, \cdots$.

Proof. Since $\mathrm{supp}\,(\mathbf{x}_\perp^\star) \subseteq \mathcal{S} = \mathrm{supp}\,(\mathbf{x}^\star)$ we have $\mathbf{d}^k|_{\mathcal{S}^c} = \mathbf{x}^k|_{\mathcal{S}^c}$. Furthermore, because $\mathbf{x}^k$ is a feasible point by assumption we have $\left\| \mathbf{x}^k \right\|_p^p \leq \widehat{c} = \left\| \mathbf{x}_\perp^\star \right\|_p^p$ that implies,

$$
\begin{aligned}
\left\| \mathbf{d}^k|_{\mathcal{S}^c} \right\|_p^p &= \left\| \mathbf{x}^k|_{\mathcal{S}^c} \right\|_p^p \\
&\leq \left\| \mathbf{x}_\perp^\star \right\|_p^p - \left\| \mathbf{x}^k|_{\mathcal{S}} \right\|_p^p \\
&\leq \left\| \mathbf{x}_\perp^\star - \mathbf{x}^k|_{\mathcal{S}} \right\|_p^p \\
&= \left\| \mathbf{d}^k|_{\mathcal{S}} \right\|_p^p \\
&\leq s^{1 - \frac{p}{2}} \left\| \mathbf{d}^k|_{\mathcal{S}} \right\|_2^p, \qquad\qquad \text{(power means inequality)}
\end{aligned}
$$

which yields the desired result. ∎

The next lemma is a straightforward extension of a previously known result [11, Lemma 3.1] to the case of complex vectors and asymmetric RIP.

Lemma 2.6. *For* $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$ *suppose that matrix* $\mathbf{A}$ *satisfies RIP of order* $\max\left(\left\| \mathbf{u} + \mathbf{v} \right\|_0, \left\| \mathbf{u} - \mathbf{v} \right\|_0 \right)$ *with constants* α *and* β. *Then we have*

$$
\left| \Re \left[\eta \left\langle \mathbf{Au}, \mathbf{Av} \right\rangle - \left\langle \mathbf{u}, \mathbf{v} \right\rangle \right] \right| \leq \left(\frac{\eta\,(\alpha - \beta)}{2} + \left| \frac{\eta\,(\alpha + \beta)}{2} - 1 \right| \right) \left\| \mathbf{u} \right\|_2 \left\| \mathbf{v} \right\|_2.
$$

Proof. If either of the vectors $\mathbf{u}$ and $\mathbf{v}$ is zero the claim becomes trivial. So without loss of generality we assume that none of these vectors is zero. The RIP condition holds for the vectors $\mathbf{u} \pm \mathbf{v}$ and we have

$$
\beta \left\| \mathbf{u} \pm \mathbf{v} \right\|_2^2 \leq \left\| \mathbf{A}\,(\mathbf{u} \pm \mathbf{v}) \right\|_2^2 \leq \alpha \left\| \mathbf{u} \pm \mathbf{v} \right\|_2^2.
$$

Therefore, we obtain

$$
\begin{aligned}
\Re \left\langle \mathbf{Au}, \mathbf{Av} \right\rangle &= \frac{1}{4} \left(\left\| \mathbf{A}\,(\mathbf{u} + \mathbf{v}) \right\|_2^2 - \left\| \mathbf{A}\,(\mathbf{u} - \mathbf{v}) \right\|_2^2 \right) \\
&\leq \frac{1}{4} \left(\alpha \left\| \mathbf{u} + \mathbf{v} \right\|_2^2 - \beta \left\| \mathbf{u} - \mathbf{v} \right\|_2^2 \right) \\
&= \frac{\alpha - \beta}{4} \left(\left\| \mathbf{u} \right\|_2^2 + \left\| \mathbf{v} \right\|_2^2 \right) + \frac{\alpha + \beta}{2} \Re \left\langle \mathbf{u}, \mathbf{v} \right\rangle.
\end{aligned}
$$

Applying this inequality for vectors $\frac{\mathbf{u}}{\left\| \mathbf{u} \right\|_2}$ and $\frac{\mathbf{v}}{\left\| \mathbf{v} \right\|_2}$ yields

$$
\begin{aligned}
\Re \left[\eta \left\langle \mathbf{A} \frac{\mathbf{u}}{\left\| \mathbf{u} \right\|_2}, \mathbf{A} \frac{\mathbf{v}}{\left\| \mathbf{v} \right\|_2} \right\rangle - \left\langle \frac{\mathbf{u}}{\left\| \mathbf{u} \right\|_2}, \frac{\mathbf{v}}{\left\| \mathbf{v} \right\|_2} \right\rangle \right] &\leq \frac{\eta\,(\alpha - \beta)}{2} + \left(\frac{\eta\,(\alpha + \beta)}{2} - 1 \right) \Re \left\langle \frac{\mathbf{u}}{\left\| \mathbf{u} \right\|_2}, \frac{\mathbf{v}}{\left\| \mathbf{v} \right\|_2} \right\rangle \\
&\leq \frac{\eta\,(\alpha - \beta)}{2} + \left| \frac{\eta\,(\alpha + \beta)}{2} - 1 \right|.
\end{aligned}
$$

Similarly it can be shown that

$$
\Re \left[\eta \left\langle \mathbf{A} \frac{\mathbf{u}}{\left\| \mathbf{u} \right\|_2}, \mathbf{A} \frac{\mathbf{v}}{\left\| \mathbf{v} \right\|_2} \right\rangle - \left\langle \frac{\mathbf{u}}{\left\| \mathbf{u} \right\|_2}, \frac{\mathbf{v}}{\left\| \mathbf{v} \right\|_2} \right\rangle \right] \geq -\frac{\eta\,(\alpha - \beta)}{2} - \left| \frac{\eta\,(\alpha + \beta)}{2} - 1 \right|.
$$

The desired result follows immediately by multiplying the last two inequalities by $\left\| \mathbf{u} \right\|_2 \left\| \mathbf{v} \right\|_2$. ∎

Lemma 2.7. *If the step-size of ℓ_p-PGD obeys $|\eta_k (\alpha_{3s} + \beta_{3s})/2 - 1| \leq \tau$ for some $\tau \geq 0$, then we have*

$$\Re\left[\left\langle \mathbf{d}^k, \mathbf{d}^{k+1} \right\rangle - \eta_k \left\langle \mathbf{A}\mathbf{d}^k, \mathbf{A}\mathbf{d}^{k+1} \right\rangle\right] \leq \left((1+\tau)\rho_{3s} + \tau\right)\left(1 + \sqrt{2p}\left(\frac{2}{2-p}\right)^{\frac{1}{2}-\frac{1}{p}}\right)^2 \left\|\mathbf{d}^k\right\|_2 \left\|\mathbf{d}^{k+1}\right\|_2.$$

Proof. Note that

$$
\begin{aligned}
\Re\left[\left\langle \mathbf{d}^k, \mathbf{d}^{k+1} \right\rangle - \eta_k \left\langle \mathbf{A}\mathbf{d}^k, \mathbf{A}\mathbf{d}^{k+1} \right\rangle\right] &= \Re\left[\left\langle \mathbf{d}^k|_{\mathfrak{T}_k}, \mathbf{d}^{k+1}|_{\mathfrak{T}_{k+1}} \right\rangle - \eta_k \left\langle \mathbf{A}\mathbf{d}^k|_{\mathfrak{T}_k}, \mathbf{A}\mathbf{d}^{k+1}|_{\mathfrak{T}_{k+1}} \right\rangle\right] \\
&\quad + \sum_{i\geq 2} \Re\left[\left\langle \mathbf{d}^k|_{\mathcal{S}_{k,i}}, \mathbf{d}^{k+1}|_{\mathfrak{T}_{k+1}} \right\rangle - \eta_k \left\langle \mathbf{A}\mathbf{d}^k|_{\mathcal{S}_{k,i}}, \mathbf{A}\mathbf{d}^{k+1}|_{\mathfrak{T}_{k+1}} \right\rangle\right] \\
&\quad + \sum_{j\geq 2} \Re\left[\left\langle \mathbf{d}^k|_{\mathfrak{T}_k}, \mathbf{d}^{k+1}|_{\mathcal{S}_{k+1,j}} \right\rangle - \eta_k \left\langle \mathbf{A}\mathbf{d}^k|_{\mathfrak{T}_k}, \mathbf{A}\mathbf{d}^{k+1}|_{\mathcal{S}_{k+1,j}} \right\rangle\right] \\
&\quad + \sum_{i,j\geq 2} \Re\left[\left\langle \mathbf{d}^k|_{\mathcal{S}_{k,i}}, \mathbf{d}^{k+1}|_{\mathcal{S}_{k+1,j}} \right\rangle - \eta_k \left\langle \mathbf{A}\mathbf{d}^k|_{\mathcal{S}_{k,i}}, \mathbf{A}\mathbf{d}^{k+1}|_{\mathcal{S}_{k+1,j}} \right\rangle\right]. \quad (15)
\end{aligned}
$$

Note that $|\mathfrak{T}_k \cup \mathfrak{T}_{k+1}| \leq 3s$. Furthermore, for $i, j \geq 2$ we have $|\mathfrak{T}_k \cup \mathcal{S}_{k+1,j}| \leq 3s$, $|\mathfrak{T}_{k+1} \cup \mathcal{S}_{k,i}| \leq 3s$, and $|\mathcal{S}_{k,i} \cup \mathcal{S}_{k+1,j}| \leq 2s$. Therefore, by applying Lemma 2.6 for each of the summands in (15) and using the fact that

$$
\begin{aligned}
\rho'_{3s} &:= (1+\tau)\rho_{3s} + \tau \\
&\geq \eta_k(\alpha_{3s} - \beta_{3s})/2 + |\eta_k(\alpha_{3s} + \beta_{3s})/2 - 1|
\end{aligned}
$$

we obtain

$$
\begin{aligned}
\Re\left[\left\langle \mathbf{d}^k, \mathbf{d}^{k+1} \right\rangle - \eta_k \left\langle \mathbf{A}\mathbf{d}^k, \mathbf{A}\mathbf{d}^{k+1} \right\rangle\right] &\leq \rho'_{3s} \left\|\mathbf{d}^k|_{\mathfrak{T}_k}\right\|_2 \left\|\mathbf{d}^{k+1}|_{\mathfrak{T}_{k+1}}\right\|_2 + \sum_{i\geq 2} \rho'_{3s} \left\|\mathbf{d}^k|_{\mathcal{S}_{k,i}}\right\|_2 \left\|\mathbf{d}^{k+1}|_{\mathfrak{T}_{k+1}}\right\|_2 \\
&\quad + \sum_{j\geq 2} \rho'_{3s} \left\|\mathbf{d}^k|_{\mathfrak{T}_k}\right\|_2 \left\|\mathbf{d}^{k+1}|_{\mathcal{S}_{k+1,j}}\right\|_2 + \sum_{i,j\geq 2} \rho'_{3s} \left\|\mathbf{d}^k|_{\mathcal{S}_{k,i}}\right\|_2 \left\|\mathbf{d}^{k+1}|_{\mathcal{S}_{k+1,j}}\right\|_2.
\end{aligned}
$$

Hence, applying Lemma 2.4 yields

$$
\begin{aligned}
\Re\left[\left\langle \mathbf{d}^k, \mathbf{d}^{k+1} \right\rangle - \eta_k \left\langle \mathbf{A}\mathbf{d}^k, \mathbf{A}\mathbf{d}^{k+1} \right\rangle\right] &\leq \rho'_{3s} \left\|\mathbf{d}^k|_{\mathfrak{T}_k}\right\|_2 \left\|\mathbf{d}^{k+1}|_{\mathfrak{T}_{k+1}}\right\|_2 \\
&\quad + \sqrt{2p}\left(\frac{2s}{2-p}\right)^{\frac{1}{2}-\frac{1}{p}} \rho'_{3s} \left\|\mathbf{d}^k|_{\mathcal{S}^c}\right\|_p \left\|\mathbf{d}^{k+1}|_{\mathfrak{T}_{k+1}}\right\|_2 \\
&\quad + \sqrt{2p}\left(\frac{2s}{2-p}\right)^{\frac{1}{2}-\frac{1}{p}} \rho'_{3s} \left\|\mathbf{d}^k|_{\mathfrak{T}_k}\right\|_2 \left\|\mathbf{d}^{k+1}|_{\mathcal{S}^c}\right\|_p \\
&\quad + 2p\left(\frac{2s}{2-p}\right)^{1-\frac{2}{p}} \rho'_{3s} \left\|\mathbf{d}^k|_{\mathcal{S}^c}\right\|_p \left\|\mathbf{d}^{k+1}|_{\mathcal{S}^c}\right\|_p.
\end{aligned}
$$

Then it follows from Lemma 2.5,

$$
\begin{aligned}
\Re\left[\left\langle \mathbf{d}^k, \mathbf{d}^{k+1} \right\rangle - \eta_k \left\langle \mathbf{A}\mathbf{d}^k, \mathbf{A}\mathbf{d}^{k+1} \right\rangle\right] &\leq \rho'_{3s} \left\|\mathbf{d}^k|_{\mathfrak{T}_k}\right\|_2 \left\|\mathbf{d}^{k+1}|_{\mathfrak{T}_{k+1}}\right\|_2 \\
&\quad + \sqrt{2p}\left(\frac{2}{2-p}\right)^{\frac{1}{2}-\frac{1}{p}} \rho'_{3s} \left\|\mathbf{d}^k|_{\mathcal{S}}\right\|_2 \left\|\mathbf{d}^{k+1}|_{\mathfrak{T}_{k+1}}\right\|_2 \\
&\quad + \sqrt{2p}\left(\frac{2}{2-p}\right)^{\frac{1}{2}-\frac{1}{p}} \rho'_{3s} \left\|\mathbf{d}^k|_{\mathfrak{T}_k}\right\|_2 \left\|\mathbf{d}^{k+1}|_{\mathcal{S}}\right\|_2 \\
&\quad + 2p\left(\frac{2}{2-p}\right)^{1-\frac{2}{p}} \rho'_{3s} \left\|\mathbf{d}^k|_{\mathcal{S}}\right\|_2 \left\|\mathbf{d}^{k+1}|_{\mathcal{S}}\right\|_2
\end{aligned}
$$

$$\leq \rho'_{3s} \left(1 + \sqrt{2p} \left(\frac{2}{2-p}\right)^{\frac{1}{2}-\frac{1}{p}}\right)^2 \left\|\mathbf{d}^k\right\|_2 \left\|\mathbf{d}^{k+1}\right\|_2$$

∎

Now we are ready to prove the accuracy guarantees for the ℓ_p-PGD algorithm.

Proof of Theorem 2.1. Recall that γ is defined by (11). It follows from Lemmas 2.1 and 2.7 that

$$\left\|\mathbf{d}^k\right\|_2^2 \leq 2\gamma \left\|\mathbf{d}^k\right\|_2 \left\|\mathbf{d}^{k-1}\right\|_2 + 2\eta_k \Re \left\langle \mathbf{A}\mathbf{d}^k, \mathbf{A}\mathbf{d}^\star + \mathbf{e}\right\rangle$$
$$\leq 2\gamma \left\|\mathbf{d}^k\right\|_2 \left\|\mathbf{d}^{k-1}\right\|_2 + 2\eta_k \left\|\mathbf{A}\mathbf{d}^k\right\|_2 \left\|\mathbf{A}\mathbf{d}^\star + \mathbf{e}\right\|_2.$$

Furthermore, using (14) and Lemma 2.5 we deduce

$$\left\|\mathbf{A}\mathbf{d}^k\right\|_2 \leq \left\|\mathbf{A}\mathbf{d}^k|_{\mathcal{T}_k}\right\|_2 + \sum_{i \geq 1} \left\|\mathbf{A}\mathbf{d}^k|_{\mathcal{T}_{k,2i}}\right\|_2$$

$$\leq \sqrt{\alpha_{2s}} \left\|\mathbf{d}^k|_{\mathcal{T}_k}\right\|_2 + \sum_{i \geq 1} \sqrt{\alpha_{2s}} \left\|\mathbf{d}^k|_{\mathcal{T}_{k,2i}}\right\|_2$$

$$\leq \sqrt{\alpha_{2s}} \left\|\mathbf{d}^k|_{\mathcal{T}_k}\right\|_2 + \sqrt{\alpha_{2s}} \sqrt{p} \left(\frac{2s}{2-p}\right)^{\frac{1}{2}-\frac{1}{p}} \sum_{i \geq 1} \left\|\mathbf{d}^k|_{\mathcal{T}_{k,2i-1}}\right\|_p$$

$$\leq \sqrt{\alpha_{2s}} \left\|\mathbf{d}^k|_{\mathcal{T}_k}\right\|_2 + \sqrt{\alpha_{2s}} \sqrt{p} \left(\frac{2s}{2-p}\right)^{\frac{1}{2}-\frac{1}{p}} \left\|\mathbf{d}^k|_{\mathcal{S}^c}\right\|_p$$

$$\leq \sqrt{\alpha_{2s}} \left\|\mathbf{d}^k|_{\mathcal{T}_k}\right\|_2 + \sqrt{\alpha_{2s}} \sqrt{p} \left(\frac{2}{2-p}\right)^{\frac{1}{2}-\frac{1}{p}} \left\|\mathbf{d}^k|_{\mathcal{S}}\right\|_2$$

$$\leq \sqrt{\alpha_{2s}} \left(1 + \sqrt{p} \left(\frac{2}{2-p}\right)^{\frac{1}{2}-\frac{1}{p}}\right) \left\|\mathbf{d}^k\right\|_2.$$

Therefore,

$$\left\|\mathbf{d}^k\right\|_2^2 \leq 2\gamma \left\|\mathbf{d}^k\right\|_2 \left\|\mathbf{d}^{k-1}\right\|_2 + 2\eta_k \sqrt{\alpha_{2s}} \left(1 + \sqrt{p} \left(\frac{2}{2-p}\right)^{\frac{1}{2}-\frac{1}{p}}\right) \left\|\mathbf{d}^k\right\|_2 \left\|\mathbf{A}\mathbf{d}^\star + \mathbf{e}\right\|_2,$$

which after canceling $\left\|\mathbf{d}^k\right\|_2$ yields

$$\left\|\mathbf{d}^k\right\|_2 \leq 2\gamma \left\|\mathbf{d}^{k-1}\right\|_2 + 2\eta_k \sqrt{\alpha_{2s}} \left(1 + \sqrt{p} \left(\frac{2}{2-p}\right)^{\frac{1}{2}-\frac{1}{p}}\right) \left\|\mathbf{A}\mathbf{d}^\star + \mathbf{e}\right\|_2$$

$$= 2\gamma \left\|\mathbf{d}^{k-1}\right\|_2 + 2\eta_k \left(\alpha_{3s} + \beta_{3s}\right) \frac{\sqrt{\alpha_{2s}}}{\alpha_{3s} + \beta_{3s}} \left(1 + \sqrt{p} \left(\frac{2}{2-p}\right)^{\frac{1}{2}-\frac{1}{p}}\right) \left\|\mathbf{A}\mathbf{d}^\star + \mathbf{e}\right\|_2$$

$$\leq 2\gamma \left\|\mathbf{d}^{k-1}\right\|_2 + 4\left(1 + \tau\right) \frac{\sqrt{\alpha_{2s}}}{\alpha_{3s} + \beta_{3s}} \left(1 + \sqrt{p} \left(\frac{2}{2-p}\right)^{\frac{1}{2}-\frac{1}{p}}\right) \left(\left\|\mathbf{A}\mathbf{d}^\star\right\|_2 + \left\|\mathbf{e}\right\|_2\right).$$

Since $\mathbf{x}_\perp^\star$ is a projection of $\mathbf{x}^\star$ onto the feasible set $\widehat{\mathcal{B}}$ and $\left(\frac{\widehat{c}}{\|\mathbf{x}^\star\|_p^p}\right)^{1/p} \mathbf{x}^\star \in \widehat{\mathcal{B}}$ we have

$$\left\|\mathbf{d}^\star\right\|_2 = \left\|\mathbf{x}_\perp^\star - \mathbf{x}^\star\right\|_2$$

$$\leq \left\|\left(\frac{\widehat{c}}{\|\mathbf{x}^\star\|_p^p}\right)^{1/p} \mathbf{x}^\star - \mathbf{x}^\star\right\|_2$$

$$= \varepsilon \left\|\mathbf{x}^\star\right\|_2.$$

11

Furthermore, $\operatorname{supp}(\mathbf{d}^\star) \subseteq \mathcal{S}$, thereby we can use RIP to obtain

$$\begin{aligned}
\|\mathbf{A}\mathbf{d}^\star\|_2 &\leq \sqrt{\alpha_s}\,\|\mathbf{d}^\star\|_2 \\
&\leq \varepsilon\sqrt{\alpha_s}\,\|\mathbf{x}^\star\|_2\,.
\end{aligned}$$

Hence,

$$\begin{aligned}
\left\|\mathbf{d}^k\right\|_2 &\leq 2\gamma\left\|\mathbf{d}^{k-1}\right\|_2 + 4\,(1+\tau)\,\frac{\sqrt{\alpha_{2s}}}{\alpha_{3s}+\beta_{3s}}\left(1+\sqrt{p}\left(\frac{2}{2-p}\right)^{\frac{1}{2}-\frac{1}{p}}\right)\left(\varepsilon\sqrt{\alpha_s}\,\|\mathbf{x}^\star\|_2 + \|\mathbf{e}\|_2\right) \\
&\leq 2\gamma\left\|\mathbf{d}^{k-1}\right\|_2 + 2\,(1+\tau)\left(1+\sqrt{p}\left(\frac{2}{2-p}\right)^{\frac{1}{2}-\frac{1}{p}}\right)\left(\varepsilon\,(1+\rho_{3s})\,\|\mathbf{x}^\star\|_2 + \frac{2\sqrt{\alpha_{2s}}}{\alpha_{3s}+\beta_{3s}}\,\|\mathbf{e}\|_2\right).
\end{aligned}$$

Applying this inequality recursively and using the fact that

$$\sum_{i=0}^{k-1}(2\gamma)^i < \sum_{i=0}^{\infty}(2\gamma)^i = \frac{1}{1-2\gamma},$$

which holds because of the assumption $\gamma < \frac{1}{2}$, we can finally deduce

$$\begin{aligned}
\left\|\mathbf{x}^k - \mathbf{x}^\star\right\|_2 &= \left\|\mathbf{d}^k - \mathbf{d}^\star\right\|_2 \\
&\leq \left\|\mathbf{d}^k\right\|_2 + \|\mathbf{d}^\star\|_2 \\
&\leq (2\gamma)^k\,\|\mathbf{x}_\perp^\star\|_2 + \frac{2\,(1+\tau)}{1-2\gamma}\,(1+\xi\,(p))\left(\varepsilon\,(1+\rho_{3s})\,\|\mathbf{x}^\star\|_2 + \frac{2\sqrt{\alpha_{2s}}}{\alpha_{3s}+\beta_{3s}}\,\|\mathbf{e}\|_2\right) + \|\mathbf{d}^\star\|_2 \\
&\leq (2\gamma)^k\,\|\mathbf{x}^\star\|_2 + \frac{2\,(1+\tau)}{1-2\gamma}\,(1+\xi\,(p))\left(\varepsilon\,(1+\rho_{3s})\,\|\mathbf{x}^\star\|_2 + \frac{2\sqrt{\alpha_{2s}}}{\alpha_{3s}+\beta_{3s}}\,\|\mathbf{e}\|_2\right) + \varepsilon\,\|\mathbf{x}^\star\|_2\,,
\end{aligned}$$

where $\xi\,(p) = \sqrt{p}\left(\frac{2}{2-p}\right)^{\frac{1}{2}-\frac{1}{p}}$ as defined in the statement of the theorem. ∎

3. Discussion

In this paper we studied the accuracy of the Projected Gradient Descent algorithm in solving sparse least squares problems where sparsity is dictated by an ℓ_p-norm constraint. Assuming that one has an algorithm that can find a projection of any given point onto ℓ_p-balls with $p \in [0,1]$, we have shown that the PGD method converges to the true signal, up to the statistical precision, at a linear rate. The convergence guarantees in this paper are obtained by requiring proper RIP conditions to hold for the measurement matrix. By varying p from zero to one, these sufficient conditions become more stringent while robustness to noise and convergence rate worsen. This behavior suggests that smaller values of p are preferable, and in fact the PGD method at $p = 0$ (i.e., the IHT algorithm) outperforms the PGD method at $p > 0$ in every aspect. These conclusions, however, are not definitive as we have merely presented sufficient conditions for accuracy of the PGD method.

Unfortunately and surprisingly, for $p \in (0,1)$ the algorithm for projection onto ℓ_p-balls is not as simple as the cases of $p = 0$ and $p = 1$, leaving practicality of the algorithm unclear for the intermediate values p. We have shown (see the Appendix) that a projection $\mathbf{x}^\perp$ of point $\mathbf{x} \in \mathbb{C}^n$ has the following properties

(i) $\left|x_i^\perp\right| \leq |x_i|$ for all $i \in [n]$ while there is at most one $i \in [n]$ such that $\left|x_i^\perp\right| < \frac{1-p}{2-p}\,|x_i|$,

(ii) $\operatorname{Arg}(x_i) = \operatorname{Arg}\left(x_i^\perp\right)$ for $i \in [n]$,

(iii) if $|x_i| > |x_j|$ for some $i, j \in [n]$ then $\left|x_i^\perp\right| \geq \left|x_j^\perp\right|$, and

(iv) there exist $\lambda \geq 0$ such that for all $i \in \text{supp}\left(\mathbf{x}^{\perp}\right)$ we have $\left|x_i^{\perp}\right|^{1-p}\left(|x_i| - \left|x_i^{\perp}\right|\right) = p\lambda$.

However, these properties are not sufficient for full characterization of a projection. One may ask that if the PGD method performs the best at $p = 0$ then why is it important at all to design a projection algorithm for $p > 0$? We believe that developing an efficient algorithm for projection onto ℓ_p-balls with $p \in (0,1)$ is an interesting problem that can provide a building block for other methods of sparse signal estimation involving the ℓ_p-norm. Furthermore, studying this problem may help to find an insight on how the complexity of these algorithms vary in terms of p.

In future work, we would like to examine the performance of more sophisticated first-order methods such as the Nesterov's optimal gradient methods [19] for ℓ_p-constrained least squares problems. Furthermore, it could be possible to extend the provided framework further to analyze ℓ_p-constrained minimization with objective functions other than the squared error. This generalized framework can be used in problems such as regression with generalized linear models that arise in statistics and machine learning.

References

[1] A. Agarwal, S. Negahban, M. Wainwright, Fast global convergence rates of gradient methods for high-dimensional statistical recovery, `arXiv:1104.4824v1 [stat.ML]` (2011).

[2] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM Journal on Imaging Sciences 2 (1) (2009) 183–202.

[3] T. Blumensath, M. E. Davies, Iterative hard thresholding for compressed sensing, Applied and Computational Harmonic Analysis 27 (3) (2009) 265–274.

[4] E. J. Candès, The restricted isometry property and its implications for compressed sensing, Comptes Rendus Mathematique 346 (9-10) (2008) 589–592.

[5] E. J. Candès, J. K. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, Communications on Pure and Applied Mathematics 59 (8) (2006) 1207–1223.

[6] E. J. Candès, T. Tao, Near optimal signal recovery from random projections: Universal encoding strategies?, IEEE Transactions on Information Theory 52 (12).

[7] R. Chartrand, Exact reconstruction of sparse signals via nonconvex minimization, IEEE Signal Processing Letters 14 (10) (2007) 707 –710.

[8] R. Chartrand, Nonconvex compressed sensing and error correction, in: Proceedings of the 32nd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), vol. 3, 2007, pp. 889–892.

[9] S. S. Chen, D. L. Donoho, M. A. Saunders, Atomic decomposition by basis pursuit, SIAM Journal on Scientific Computing 20 (1) (1998) 33–61.

[10] W. Dai, O. Milenkovic, Subspace pursuit for compressive sensing signal reconstruction, IEEE Transactions on Information Theory 55 (5) (2009) 2230–2249.

[11] M. Davenport, M. Wakin, Analysis of orthogonal matching pursuit using the restricted isometry property, IEEE Transactions on Information Theory 56 (9) (2010) 4395 –4401.

[12] D. Donoho, M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization, Proceedings of the National Academy of Sciences of the United States of America 100 (5) (2003) 2197.

[13] S. Foucart, Sparse recovery algorithms: sufficient conditions in terms of restricted isometry constants, in: Approximation Theory XIII: San Antonio 2010, vol. 13 of Springer Proceedings in Mathematics, Springer New York, 2012.

[14] S. Foucart, M.-J. Lai, Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$, Applied and Computational Harmonic Analysis 26 (3) (2009) 395–407.

[15] R. Gribonval, M. Nielsen, Highly sparse representations from dictionaries are unique and independent of the sparseness measure, Applied and Computational Harmonic Analysis 22 (3) (2007) 335–355.

[16] A. Maleki, D. Donoho, Optimally tuned iterative reconstruction algorithms for compressed sensing, Selected Topics in Signal Processing, IEEE Journal of 4 (2) (2010) 330 –341.

[17] B. Natarajan, Sparse approximate solutions to linear systems, SIAM journal on computing 24 (1995) 227.

[18] D. Needell, J. A. Tropp, CoSaMP: iterative signal recovery from incomplete and inaccurate samples, Applied and Computational Harmonic Analysis 26 (3) (2009) 301–321.

[19] Y. Nesterov, Introductory Lectures on Convex Optimization: A Basic Course, Kluwer Academic Publishers, 2004.

[20] R. Saab, R. Chartrand, Ö. Yilmaz, Stable sparse approximations via nonconvex optimization, in: Proceedings of the 33rd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2008, pp. 3885–3888.

[21] R. Saab, Ö. Yilmaz, Sparse recovery by non-convex optimization - instance optimality, Applied and Computational Harmonic Analysis 29 (1) (2010) 30–48.

[22] R. Tibshirani, Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society. Series B (Methodological) 58 (1) (1996) 267–288.

Appendix A. Lemmas for Characterization of a Projection onto ℓ_p-balls

In what follows we assume that $\mathcal{B}$ is an ℓ_p-ball with p-radius c (i.e., $\mathcal{B} = \mathcal{F}_p(c)$). For $\mathbf{x} \in \mathbb{C}^n$ we derive some properties of

$$\mathbf{x}^\perp \in \arg\min \frac{1}{2} \|\mathbf{x} - \mathbf{u}\|_2^2 \quad \text{s.t. } \mathbf{u} \in \mathcal{B}, \tag{A.1}$$

a projection of $\mathbf{x}$ onto $\mathcal{B}$.

Lemma A.1. *Let $\mathbf{x}^\perp$ be a projection of $\mathbf{x}$ onto $\mathcal{B}$. Then for every $i \in \{1, 2, \ldots, n\}$ we have $\mathrm{Arg}(x_i) = \mathrm{Arg}(x_i^\perp)$ and $|x_i^\perp| \leq |x_i|$.*

Proof. Proof by contradiction. Suppose that for some i we have $\mathrm{Arg}(x_i) \neq \mathrm{Arg}(x_i^\perp)$ or $|x_i^\perp| > |x_i|$. Consider the vector $\mathbf{x}'$ for which $x_j' = x_j^\perp$ for $j \neq i$ and $x_i' = \min\{|x_i|, |x_i^\perp|\} \exp(\imath\mathrm{Arg}(x_i))$ (the character $\imath$ denotes the imaginary unit $\sqrt{-1}$). We have $\|\mathbf{x}'\|_p \leq \|\mathbf{x}^\perp\|_p$ which implies that $\mathbf{x}' \in \mathcal{B}$. Since $|x_i - x_i'| < |x_i - x_i^\perp|$ we have $\|\mathbf{x}' - \mathbf{x}\|_2 < \|\mathbf{x}^\perp - \mathbf{x}\|_2$ which contradicts the choice of $\mathbf{x}^\perp$ as a projection. ∎

Assumption. *Lemma A.1 asserts that the projection $\mathbf{x}^\perp$ has the same phase components as $\mathbf{x}$. Therefore, without loss of generality and for simplicity in the following lemmas we assume $\mathbf{x}$ has real-valued non-negative entries.*

Lemma A.2. *For any $\mathbf{x}$ in the positive orthant there is a projection $\mathbf{x}^\perp$ of $\mathbf{x}$ onto the set $\mathcal{B}$ such that for $i, j \in \{1, 2, \ldots, n\}$ we have $x_i^\perp \leq x_j^\perp$ iff $x_i \leq x_j$.*

Proof. Note that the set $\mathcal{B}$ is closed under any permutation of coordinates. In particular, by interchanging the i-th and j-th entries of $\mathbf{x}^\perp$ we obtain another vector $\mathbf{x}'$ in $\mathcal{B}$. Since $\mathbf{x}^\perp$ is a projection of $\mathbf{x}$ onto $\mathcal{B}$ we must have $\|\mathbf{x} - \mathbf{x}^\perp\|_2^2 \leq \|\mathbf{x} - \mathbf{x}'\|_2^2$. Therefore, we have $(x_i - x_i^\perp)^2 + (x_j - x_j^\perp)^2 \leq (x_i - x_j^\perp)^2 + (x_j - x_i^\perp)^2$ and from that $0 \leq (x_i - x_j)(x_i^\perp - x_j^\perp)$. For $x_i \neq x_j$ the result follows immediately, and for $x_i = x_j$ without loss of generality we can assume $x_i^\perp \leq x_j^\perp$. ∎

Lemma A.3. *Let $\mathcal{S}^\perp$ be the support set of $\mathbf{x}^\perp$. Then there exists a $\lambda \geq 0$ such that*

$$x_i^{\perp(1-p)}(x_i - x_i^\perp) = p\lambda$$

for all $i \in \mathcal{S}^\perp$.

Proof. The fact that $\mathbf{x}^\perp$ is a solution to the minimization expressed in (A.1) implies that that $\mathbf{x}^\perp|_{\mathcal{S}^\perp}$ must be a solution to

$$\arg\min_{\mathbf{v}} \frac{1}{2}\|\mathbf{x}|_{\mathcal{S}^\perp} - \mathbf{v}\|_2^2 \quad \text{s.t. } \|\mathbf{v}\|_p^p \leq c.$$

The normal to the feasible set (i.e., the gradient of the constraint function) is uniquely defined at $\mathbf{x}^\perp|_{\mathcal{S}^\perp}$ since all of its entries are positive by assumption. Consequently, the Lagrangian

$$L(\mathbf{v}, \lambda) = \frac{1}{2}\|\mathbf{x}|_{\mathcal{S}^\perp} - \mathbf{v}\|_2^2 + \lambda\left(\|\mathbf{v}\|_p^p - c\right)$$

has a well-defined partial derivative $\frac{\partial L}{\partial \mathbf{v}}$ at $\mathbf{x}^\perp|_{\mathcal{S}^\perp}$ which must be equal to zero for an appropriate $\lambda \geq 0$. Hence,

$$\forall i \in \mathcal{S}^\perp \quad x_i^\perp - x_i + p\lambda x_i^{\perp(p-1)} = 0$$

which is equivalent to the desired result. ∎

Lemma A.4. *Let $\lambda \geq 0$ and $p \in [0,1]$ be fixed numbers and set $T_0 = (2-p)\left(p(1-p)^{p-1}\lambda\right)^{\frac{1}{2-p}}$. Denote the function $t^{1-p}(T-t)$ by $h_p(t)$. The following statements hold regarding the roots of $h_p(t) = p\lambda$:*

(i) *For $p = 1$ and $T \geq T_0$ the equation $h_1(t) = \lambda$ has a unique solution at $t = T - \lambda \in [0,T]$ which is an increasing function of T.*

(ii) *For $p \in [0,1)$ and $T \geq T_0$ the equation $h_p(t) = p\lambda$ has two roots t_- and t_+ satisfying $t_- \in \left(0, \frac{1-p}{2-p}T\right]$ and $t_+ \in \left[\frac{1-p}{2-p}T, +\infty\right)$. As a function of T, t_- and t_+ are decreasing and increasing, respectively and they coincide at $T = T_0$.*

Proof. Fig. A.3 illustrates $h_p(t)$ for different values of $p \in [0,1]$. To verify part (i) observe that we have $T_0 = \lambda$ thereby $T \geq \lambda$. The claim is then obvious since $h_1(t) - \lambda = T - t - \lambda$ is zero at $t = T - \lambda$. Part (ii) is more intricate and we divide it into two cases: $p = 0$ and $p \neq 0$. At $p = 0$ we have $T_0 = 0$ and $h_0(t) = t(T-t)$ has two zeros at $t_- = 0$ and $t_+ = T$ that obviously satisfy the claim. So we can now focus on the case $p \in (0,1)$. It is straightforward to verify that $t_{\max} = \frac{1-p}{2-p}T$ is the location at which $h_p(t)$ peaks. Straightforward algebraic manipulations also show that $T > T_0$ is equivalent to $p\lambda < h_p(t_{\max})$. Furthermore, inspecting the sign of $h_p'(t)$ shows that $h_p(t)$ is strictly increasing over $[0, t_{\max}]$ while it is strictly decreasing over $[t_{\max}, T]$. Then, using the fact that $h_p(0) = h_p(T) = 0 \leq p\lambda < h_p(t_{\max})$, it follows from the *intermediate value theorem* that $h_p(t) = p\lambda$ has exactly two roots, t_- and t_+, that straddle $t_{\max}$ as claimed. Furthermore, taking the derivative of $t_-^{1-p}(T - t_-) = p\lambda$ with respect to T yields

$$(1-p)\, t_-'\, t_-^{-p}(T - t_-) + t_-^{1-p}(1 - t_-') = 0.$$

Hence,

$$((1-p)(T - t_-) - t_-)\, t_-' = -t_-$$

which because $t_- \leq t_{\max} = \frac{1-p}{2-p}T$ implies that $t_-' < 0$. Thus t_- is a decreasing function of T. Similarly we can show that t_+ is an increasing function of T using the fact that $t_+ \geq t_{\max}$. Finally, as T decreases to T_0 the peak value $h_p(t_{\max})$ decreases to $p\lambda$ which implies that t_- and t_+ both tend to the same value of $\frac{1-p}{2-p}T_0$. ∎

Lemma A.5. *Suppose that $x_i = x_j > 0$ for some $i \neq j$. If $x_i^\perp = x_j^\perp > 0$ then $x_i^\perp \geq \frac{1-p}{2-p}x_i$.*

Proof. For $p \in \{0,1\}$ the claim is obvious since at $p = 0$ we have $x_i^\perp = x_i > \frac{1}{2}x_i$ and at $p = 1$ we have $\frac{1-p}{2-p}x_i = 0$. Therefore, without loss of generality we assume $p \in (0,1)$. The proof is by contradiction. Suppose that $w = \frac{x_i^\perp}{x_i} = \frac{x_j^\perp}{x_j} < \frac{1-p}{2-p}$. Since $\mathbf{x}^\perp$ is a projection it follows that $a = b = w$ must be the solution to

$$\arg\min_{a,b}\ \psi = \frac{1}{2}\left[(1-a)^2 + (1-b)^2\right] \quad \text{s.t. } a^p + b^p = 2w^p,\ a > 0,\ \text{and } b > 0,$$

otherwise the vector $\mathbf{x}'$ that is identical to $\mathbf{x}^\perp$ except for $x_i' = ax_i \neq x_i^\perp$ and $x_j' = bx_j \neq x_i^\perp$ is also a feasible point (i.e., $\mathbf{x}' \in \mathcal{B}$) that satisfies

$$\|\mathbf{x}' - \mathbf{x}\|_2^2 - \|\mathbf{x}^\perp - \mathbf{x}\|_2^2 = (1-a)^2 x_i^2 + (1-b)^2 x_j^2 - (1-w)^2 x_i^2 - (1-w)^2 x_j^2$$
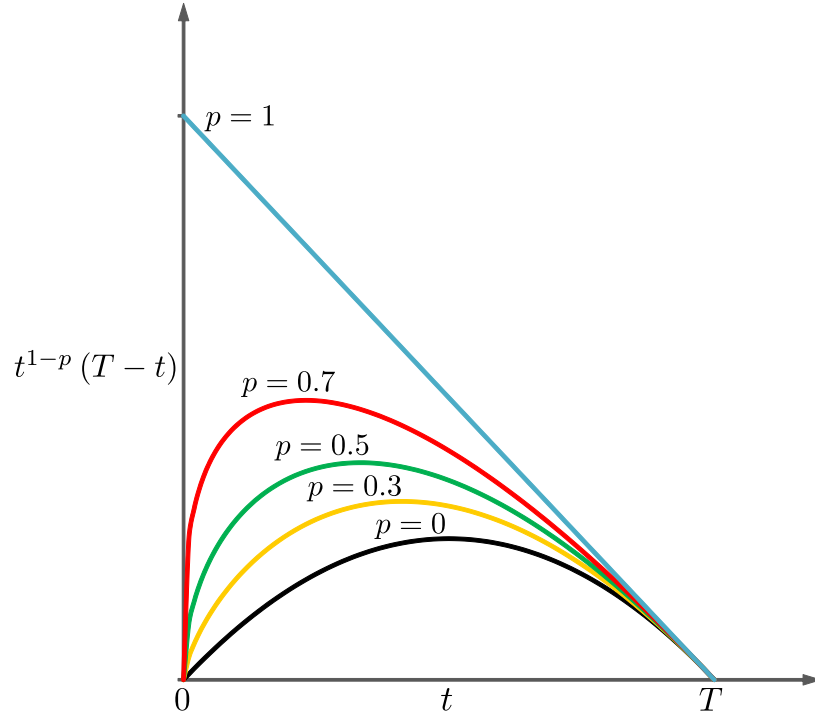$$= \left((1-a)^2 + (1-b)^2 - 2(1-w)^2\right)x_i^2$$
$$< 0,$$

Figure A.3: The function $t^{1-p}(T-t)$ for different values of p

which is absurd. If b is considered as a function of a then ψ can be seen merely as a function of a, i.e., $\psi \equiv \psi(a)$. Taking the derivative of ψ with respect to a yields

$$\psi'(a) = a - 1 + b'(b-1)$$

$$= a - 1 - \left(\frac{a}{b}\right)^{p-1}(b-1)$$

$$= \left(b^{1-p}(1-b) - a^{1-p}(1-a)\right)a^{p-1}$$

$$= (2-p)(b-a)\nu^{-p}\left(\frac{1-p}{2-p} - \nu\right),$$

where the last equation holds by the *mean value theorem* for some $\nu \in (\min\{a,b\}, \max\{a,b\})$. Since $w < \frac{1-p}{2-p}$ we have $r_1 := \min\left\{2^{1/p}w, \frac{1-p}{2-p}\right\} > w$ and $r_0 := (2w^p - r_1^p)^{1/p} < w$. With straightforward algebra one can show that if either a or b belongs to the interval $[r_0, r_1]$, then so does the other one. By varying a in $[r_0, r_1]$ we always have $\nu < r_1 \leq \frac{1-p}{2-p}$, therefore as a increases in this interval the sign of ψ' changes at $a = w$ from positive to negative. Thus, $a = b = w$ is a local maximum of ψ which is a contradiction. ∎